

## Automation using Python

Automation script which accept directory name from user and display all names of duplicate files from that directory.

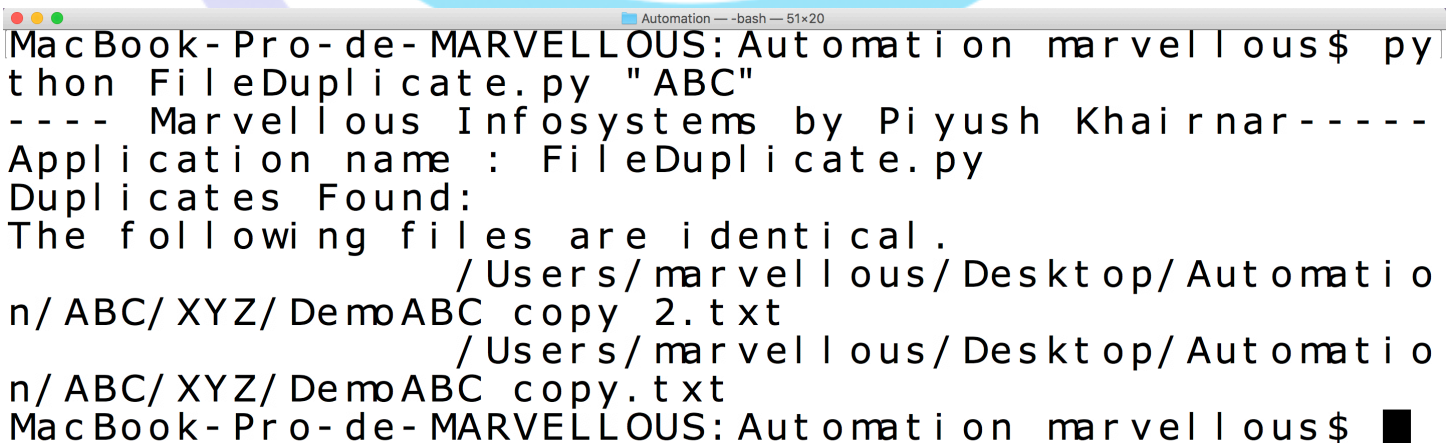
```
1 from sys import *
2 import os
3 import hashlib
4
5 def hashfile(path, blocksize = 1024):
6     fd = open(path, 'rb')
7     hasher = hashlib.md5()
8     buf = fd.read(blocksize)
9
10    while len(buf) > 0:
11        hasher.update(buf)
12        buf = fd.read(blocksize)
13
14    fd.close()
15
16    return hasher.hexdigest()
17
18 def FindDuplicate(path):
19     flag = os.path.isabs(path)
20
21     if flag == False:
22         path = os.path.abspath(path)
23
24     exists = os.path.isdir(path)
25
26     dups = {}
27
28     if exists:
29         for dirName, subdirs, fileList in os.walk(path):
30             for filen in fileList:
31                 path = os.path.join(dirName, filen)
32                 file_hash = hashfile(path)
33                 if file_hash in dups:
34                     dups[file_hash].append(path)
35                 else:
36                     dups[file_hash] = [path]
37
38     return dups
39 else:
40     print("Invalid Path")
41
42 def PrintDuplicate(dict1):
43     results = list(filter(lambda x: len(x) > 1, dict1.values()))
44
45     if len(results) > 0:
46         print("Duplicates Found:")
47
48         print("The following files are identical.")
49
50         icnt = 0;
51         for result in results:
52             for subresult in result:
53                 icnt+=1
54                 if icnt >= 2:
55                     print('\t\t%s' % subresult)
```

```

57 else:
58     print("No duplicate files found.")
59
60 def main():
61     print("---- Marvellous Infosystems by Piyush Khairnar----")
62
63     print("Application name : " +argv[0])
64
65     if (len(argv) != 2):
66         print("Error : Invalid number of arguments")
67         exit()
68
69     if (argv[1] == "-h") or (argv[1] == "-H"):
70         print("This Script is used to traverse specific directory and display sizes of files")
71         exit()
72
73     if (argv[1] == "-u") or (argv[1] == "-U"):
74         print("usage : ApplicationName AbsolutePath_of_Directory Extention")
75         exit()
76
77     try:
78         arr = {}
79         arr = FindDuplicate(argv[1])
80         PrintDuplicate(arr)
81
82     except ValueError:
83         print("Error : Invalid datatype of input")
84
85 if __name__ == "__main__":
86     main()
87
88
89
90

```

## Output of above script



```

MacBook-Pro-de-MARVELLOUS:Automation marvellous$ python FileDuplicate.py "ABC"
---- Marvellous Infosystems by Piyush Khairnar----
Application name : FileDuplicate.py
Duplicates Found:
The following files are identical.
/Users/marvellous/Desktop/Automation/ABC/XYZ/DemoABC copy 2.txt
/Users/marvellous/Desktop/Automation/ABC/XYZ/DemoABC copy.txt
MacBook-Pro-de-MARVELLOUS:Automation marvellous$

```